Robust global image registration based on a hybrid algorithm combining Fourier and spatial domain techniques

Peter N. Crabtree, Collin Seanor, Jeremy Murray-Krezan, and Patrick J. McNicholl
Air Force Research Laboratory, Space Vehicles Directorate
3550 Aberdeen Ave. SE, Kirtland AFB, NM USA

ABSTRACT

A variety of image registration techniques have been investigated for applications such as image analysis, fusion, compression, enhancement, and creating mosaics. In particular, robust registration is a key component of successful multi-frame processing aimed at super-resolution or high dynamic range imaging of space objects. Image registration techniques are broadly categorized as global (area) or feature-based, and can also be classified as being performed in either the Fourier or spatial (image) domain. Spatial domain methods are typically used for applications requiring accurate estimation of sub-pixel motion, such as multi-frame super-resolution based on de-aliasing. However, these techniques often rely on the availability of *a priori* information (good initial guess), and are therefore limited in terms of dynamic range of the global relative motions between camera and scene. A Gaussian pyramid approach is one standard method to extend the region of convergence of spatial domain techniques. On the other hand, Fourier domain-based correlation techniques such as the log-polar FFT (L-P FFT) method provide fast and reasonably accurate estimates of global shifts, rotation, and uniform scale change, and tend to perform well over a large range of frame-to-frame motion magnitudes. In this paper we begin to explore possible hybrid algorithms for robust global registration based on combining the L-P FFT and spatial domain techniques. Initial results are presented for a hybrid algorithm using the output of the L-P FFT method as an initial guess for a spatial domain technique. In addition, we explore the potential benefits of normalized gradient correlation (NGC) in performing the coarse L-P FFT registration. The use of NGC, as opposed to phase-only correlation, has recently been explored for improving performance of the L-P FFT method in terms of robustness and dynamic range of scale factor estimates. The results presented here are based on both simulated image sequences and image sets captured in the laboratory using a CMOS machine vision camera mounted on computer-controlled micro-stepping motion stages. We have also begun to investigate optical flow algorithms for application to image enhancement in scenarios where relative motions cannot be adequately described by affine transformations. Initial results from our optical flow algorithm studies are presented based on well-controlled image sequences captured in the laboratory.

1. INTRODUCTION

This paper presents results from an investigation of image registration and optical flow techniques. Algorithm performance evaluations were accomplished using both simulated image sequences and image sets collected in the laboratory. Laboratory images were captured using a visible machine vision camera and well-controlled object motions using one or more computer-controlled micro-stepping motion stages. Section 2 provides background on global image registration using both Fourier domain and iterative spatial domain techniques. Methodology and approach are described in Sec. 3. Section 4 presents results for registration algorithm performance based on several simulated image sequences, as well as laboratory imagery of an ISO 12233 resolution chart. Early results from our investigation of optical flow are also presented, based on laboratory imagery of a spinning metal box. Section 5 provides a summary of results and conclusions, and outlines plans for future work.

2. BACKGROUND

Image registration is a key component in a variety of computer vision applications. Broad overviews of image registration are provided by Brown [1] and Zitová and Flusser [2]. In particular, the log-polar FFT (L-P FFT) algorithm is one well-known technique able to estimate global x- and y-axis translations, rotation, and uniform scale change as described by the following affine coordinate transformation between two images f_1 and f_2:

$$f_2(x,y) = f_1\left(xs\cos\theta_0 + ys\sin\theta_0 + t_x\,, -xs\sin\theta_0 + ys\cos\theta_0 + t_y\right) \quad , \tag{1}$$

where t_x and t_y represent translations along the x- and y-axis, respectively, θ_0 is rotation, and s is the uniform scale (zoom) factor applied to both axes. The L-P FFT algorithm utilizes three separate correlations to estimate t_x, t_y, θ_0,

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEP 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Robust global image registration based on a hybrid algorithm combining Fourier and spatial domain techniques	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory,Space Vehicles Directorate,3550 Aberdeen Ave. SE,Kirtland AFB,NM,87117**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
In Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), 11-14 Sep 2012, Maui, HI.

14. ABSTRACT

A variety of image registration techniques have been investigated for applications such as image analysis, fusion compression, enhancement, and creating mosaics. In particular, robust registration is a key component of successful multi-frame processing aimed at super-resolution or high dynamic range imaging of space objects. Image registration techniques are broadly categorized as global (area) or feature-based, and can also be classified as being performed in either the Fourier or spatial (image) domain. Spatial domain methods are typically used for applications requiring accurate estimation of sub-pixel motion, such as multi-frame super-resolution based on dealiasing. However, these techniques often rely on the availability of a priori information (good initial guess), and are therefore limited in terms of dynamic range of the global relative motions between camera and scene. A Gaussian pyramid approach is one standard method to extend the region of convergence of spatial domain techniques. On the other hand, Fourier domain-based correlation techniques such as the log-polar FFT (L-P FFT) method provide fast and reasonably accurate estimates of global shifts, rotation, and uniform scale change, and tend to perform well over a large range of frame-to-frame motion magnitudes. In this paper we begin to explore possible hybrid algorithms for robust global registration based on combining the L-P FFT and spatial domain techniques. Initial results are presented for a hybrid algorithm using the output of the L-P FFT method as an initial guess for a spatial domain technique. In addition, we explore the potential benefits of normalized gradient correlation (NGC) in performing the coarse L-P FFT registration. The use of NGC, as opposed to phase-only correlation, has recently been explored for improving performance of the L-P FFT method in terms of robustness and dynamic range of scale factor estimates. The results presented here are based on both simulated image sequences and image sets captured in the laboratory using a CMOS machine vision camera mounted on computer-controlled micro-stepping motion stages. We have also begun to investigate optical flow algorithms for application to image enhancement in scenarios where relative motions cannot be adequately described by affine transformations. Initial results from our optical flow algorithm studies are presented based on well-controlled image sequences captured in the laboratory.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**13**	

and *s*: one correlation in the Fourier domain and two in the image domain. Key steps in the L-P FFT algorithm are outlined below:

- Compute FFT of each image: $F_1 = \mathcal{F}\{f_1\}$ and $F_2 = \mathcal{F}\{f_2\}$, where $\mathcal{F}\{\cdot\}$ is the Fourier Transform operator
- Retain only the Fourier magnitudes: $M_1 = |F_1|$ and $M_2 = |F_2|$ (note that M_1 and M_2 are translation invariant)
- Map the Fourier magnitudes from Cartesian to log-polar coordinates using bilinear interpolation
- Compute the global correlation between $M_1(\log \rho, \theta)$ and $M_2(\log \rho - \log s, \theta - \theta_0)$
- Find peak in the correlation surface, the location of which provides estimates of rotation $\hat{\theta}_0$ and scale $\hat{s}$
- Create two rotated and scaled versions of the "zoomed" image, using $(\hat{s}, \hat{\theta}_0)$ and $(\hat{s}, \hat{\theta}_0 + 180°)$ (note the 180° ambiguity is due to the conjugate symmetry of the Fourier transform of a real image)
- Compute global correlations between the baseline image and the two scaled and rotated images given the two possible solutions for rotation angle
- Find the peak in each of the two correlation surfaces, the larger peak resolves the angle ambiguity and its location provides estimates of translation $\hat{t}_x$ and $\hat{t}_y$

As originally described by Reddy and Chatterji [3] the L-P FFT algorithm is based on use of phase-only correlation (PC). The more recent works of Tzimiropoulos, et al. [4, 5] recommend normalized gradient correlation (NGC) to improve L-P FFT algorithm performance as compared to PC, gradient correlation (GC), and orientation correlation (OC). We explored all four of these correlation types, but focused mainly on PC and NGC. The phase-only correlation between two images is calculated as follows:

$$\mathrm{PC} = \mathcal{F}^{-1}\{F_1 F_2^* / (|F_1||F_2|)\} \quad , \tag{2}$$

where F_1 and F_2 are the image Fourier transforms as described above. For the ideal case where two identical images are related solely by translation, the PC surface is a delta function whose location corresponds to the shift in image-space (as given by the Fourier shift theorem). The gradient-based correlations work on gray-level edge maps as opposed to the original images. The complex gradient map g_1 of image f_1 is given by

$$g_1(x,y) = \frac{\partial}{\partial x} f_1(x,y) + j \frac{\partial}{\partial y} f_1(x,y) \quad . \tag{3}$$

The complex gradient map is calculated for each original image, and the gradient correlation (GC) is then given by

$$\mathrm{GC} = \mathcal{F}^{-1}\{G_1 G_2^*\} \quad , \tag{4}$$

where G_1 and G_2 are the Fourier transforms of g_1 and g_2, respectively. The normalized gradient correlation (NGC) is now given by

$$\mathrm{NGC} = \mathrm{GC} / \mathcal{F}^{-1}\{\mathcal{F}\{|g_1|\}\mathcal{F}\{|g_2|\}^*\} \quad , \tag{5}$$

where the normalization ensures that $0 \leq |\mathrm{NGC}| \leq 1$.

The L-P FFT algorithm as described in [3-5] provides motion estimates to only pixel-level accuracy, and therefore is unsuitable in its basic form for application to multi-frame image enhancement. The L-P FFT algorithm can be directly modified using Fourier interpolation to increase resolution of the correlation peak location estimates, and thereby improve resolution of the motion estimates. We previously explored this approach [6] using the memory efficient implementation of Fourier-interpolated PC developed by Guizar-Sicairos, et al. [7]. However, successful multi-frame image enhancement algorithms [8] are more typically based on iterative spatial domain methods. Two variations of iterative image domain (i.e., forward model) registration described in the literature provide registration against translation and rotation (but not scale), and are based on Taylor series expansions of 3-parameter [9] and 4-parameter [10] affine models. The technique of Keren [9] was proposed for multi-frame image enhancement as it was found to provide robust and accurate estimation of sub-pixel motion. However, this technique is limited in terms of dynamic range of the global motion estimates, and therefore may rely on the availability of *a priori* information (good initial guess). This fact motivated our exploration of hybrid algorithms combining the log-polar FFT method for a course estimate over a large dynamic range with an iterative image domain method for improved accuracy and precision of sub-pixel motion estimates. We have initially explored the straightforward use of the log-polar FFT algorithm to generate an initial guess for use by a spatial domain algorithm, but other variations are possible based on intertwining the two methods by applying both global correlation and iterative registration at each relevant step within the log-polar FFT algorithm.

The iterative spatial domain methods of Keren [9] and Fan [10] are based on a Taylor series expansion of the translated and rotated image. The method of Keren assumes a 3-parameter affine transformation, resulting in the following approximation:

$$f_2(x,y) \approx f_1(x,y) + (t_x - y\theta_0 - x\,\theta_0^2/2)\frac{\partial f_1}{\partial x} + \left(t_y + x\theta_0 - y\,\theta_0^2/2\right)\frac{\partial f_1}{\partial y} \quad . \tag{6}$$

The sum squared error, $E(t_x, t_y, \theta_0)$, between the two images can be approximated using Eq. (6). Solving for the minimum of E, and ignoring higher-order terms, results in the following system of linear equations:

$$\begin{bmatrix} \sum\left(\frac{\partial f_1}{\partial x}\right)^2 & \sum\frac{\partial f_1}{\partial x}\frac{\partial f_1}{\partial y} & \sum\frac{\partial f_1}{\partial x}\left(x\frac{\partial f_1}{\partial y} - y\frac{\partial f_1}{\partial x}\right) \\ \sum\frac{\partial f_1}{\partial x}\frac{\partial f_1}{\partial y} & \sum\left(\frac{\partial f_1}{\partial y}\right)^2 & \sum\frac{\partial f_1}{\partial y}\left(x\frac{\partial f_1}{\partial y} - y\frac{\partial f_1}{\partial x}\right) \\ \sum\frac{\partial f_1}{\partial x}\left(x\frac{\partial f_1}{\partial y} - y\frac{\partial f_1}{\partial x}\right) & \sum\frac{\partial f_1}{\partial y}\left(x\frac{\partial f_1}{\partial y} - y\frac{\partial f_1}{\partial x}\right) & \sum\left(x\frac{\partial f_1}{\partial y} - y\frac{\partial f_1}{\partial x}\right)^2 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ \theta_0 \end{bmatrix} = \begin{bmatrix} \sum\frac{\partial f_1}{\partial x}(f_1 - f_2) \\ \sum\frac{\partial f_1}{\partial y}(f_1 - f_2) \\ \sum\left(x\frac{\partial f_1}{\partial y} - y\frac{\partial f_1}{\partial x}\right)(f_1 - f_2) \end{bmatrix} . \tag{7}$$

The solution of Eq. 7 is only valid for small values of t_x, t_y, and θ_0. Therefore, an iterative approach is recommended by Keren whereby Eq. 7 is solved, the shifted and rotated image is corrected based on this initial solution, and the process is then repeated until convergence occurs (typically 5 iterations or less based on our experience). The final solution is then given by the accumulation of the incremental estimates for t_x, t_y, and θ_0 produced by each solution of Eq. 7. The work of Fan simply extends the work of Keren by using a 4-parameter affine model instead of a 3-parameter model. This modification avoids use of the small angle approximation at one point in the derivation resulting in a somewhat increased region of convergence.

3. METHODOLOGY AND APPROACH

Our approach to developing a robust global registration method is based on a hybrid algorithm combing the (non-iterative) L-P FFT and (iterative) spatial domain techniques. This includes consideration of various ways to combine these two algorithmic methods, but our work also considers choice of correlation type at each relevant step in the L-P FFT algorithm as a potential means to increase both overall robustness and resolution of the motion estimates. Here we consider phase-only correlation (PC), gradient correlation (GC), orientation correlation (OC), and normalized gradient correlation (NGC), as well as the memory-efficient Fourier interpolated version of PC developed by Guizar-Sicairos, et al. [7].

Our overall approach to algorithm testing is to not rely solely on simulated image sequences. Therefore, we assembled a laboratory setup [11] to allow capturing of image sequences with highly-controlled motions using several computer-controlled micro-stepping motion stages available from Zaber Technologies, Inc. Our laboratory includes several visible PixeLINK CMOS machine vision cameras and an LWIR microbolometer camera. All results reported in this paper using laboratory image sequences are for visible imagery captured with a PixeLINK camera. This work considers only a single motion type in a given image sequence. In other words, translation along a given axis, rotation, and scale, are tested separately. Consideration of more complicated motions encompassing various combinations of translations, rotation, and scale change will be presented in the future, including their impact on affine parameter estimation errors. Also, the focus here has been to examine motion estimation performance of a hybrid algorithm and its components for the case of imagery corrupted by various levels of additive white Gaussian noise (AWGN), while lighting conditions are assumed static.

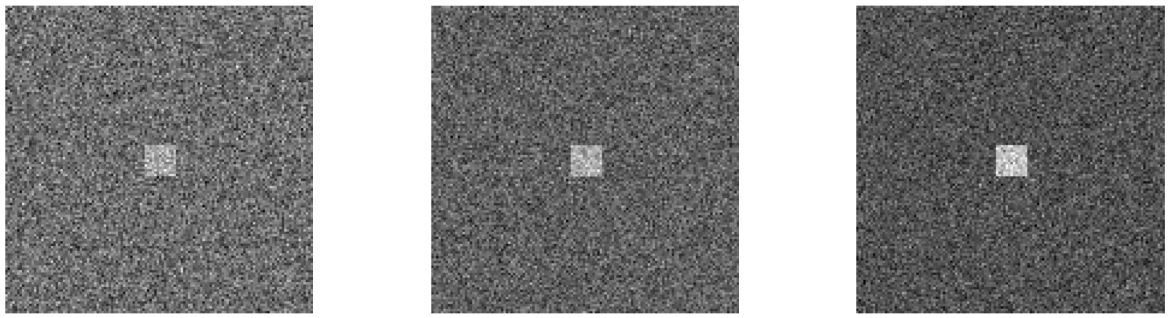
4. RESULTS

This section presents results from three distinct, but related, studies. First, results are presented from an investigation of the performance of four correlation types (PC, GC, NGC, and OC) as applied to the L-P FFT algorithm. This includes three separate tests for estimation of translation, rotation, or scale based on simulated image sequences corrupted by various levels of AWGN. Translation and scale estimation results are based on synthetic image sequences of a simple square. Rotation estimation results are for a sequence of Lenna images (cropped to 128×128 pixels) with simulated rotations using bilinear interpolation. Recall that within the L-P FFT algorithm, translation estimation entails correlation in the image domain, while rotation and scale estimation rely on correlation in the log-polar Fourier magnitude domain. Next, translation estimation results are presented to demonstrate performance of a hybrid algorithm. These results are from analysis of a set of images of an ISO 12233 [12] resolution chart captured in the laboratory using a PixeLINK camera. Finally, we present results from our optical flow investigations based on a sequence of images of a spinning metal box – also captured in the laboratory using the same PixeLINK camera.

Note that in order to achieve expected performance on simple noise-free images, we found it necessary to apply a threshold to the gradient values in all of the gradient-based correlation schemes including GC, NGC, and OC. The paper of Tzimiropoulos [5] describes a threshold for OC, but does not discuss thresholds in regards to GC or NGC. The impact of changing the threshold implementation (adaptive or static) and varying threshold level requires further study. Also, the 2nd-order central difference estimator given by kernel h = [1/12 -2/3 0 2/3 -1/12] was used to compute discrete approximations to image derivatives required for both gradient-based correlations and the iterative spatial domain optimization techniques.

TRANSLATION ESTIMATION PERFORMANCE FOR SEVERAL TYPES OF CORRELATION

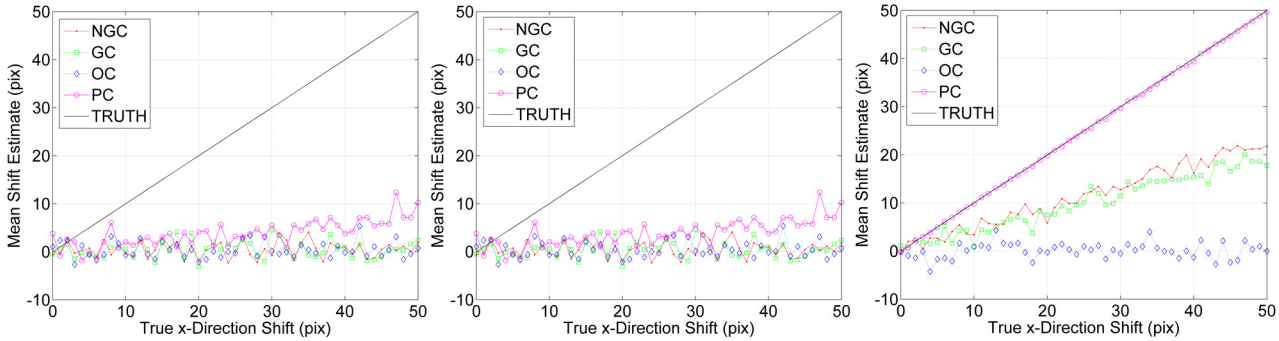
The first result is a comparison of NGC, GC, OC, and PC for the simple case of a synthetic image of a square corrupted by AWGN. The baseline test image is a 13×13 pixel matrix of ones centered within a 128×128 pixel matrix of zeros. This image was then shifted along the positive x-axis direction by an integer number of pixels ranging from t_x = [1,…,50] resulting in a 51 image sequence. Three example images are shown below in Fig. 1 for t_x = 0 and peak image SNR values of 2, 3, and 5.



Peak Image SNR = 2 Peak Image SNR = 3 Peak Image SNR = 5

Fig. 1. Synthetic test image of a 13×13 pixel matrix of ones for three different noise levels

Translation between the baseline image and each shifted image was calculated using each of the four correlation types, for 500 realizations of noise, and for peak image SNR levels of [1,2,3,4,,5,10,20,40,50,80,∞]. The expected value of the shift estimates were calculated as a function of true shift, and these results are plotted in Fig. 2 for peak image SNR values of 2, 3, and 5. The results of Fig. 2 indicate that PC clearly outperforms the other three correlation types in this simple test case. For a peak image SNR of 3 the PC correlation algorithm is beginning to properly estimate the trend of increasing shift magnitudes in the positive x-axis direction, and then for a peak SNR of 5 is providing excellent performance in terms of the expected values of the shifts. Also note that at peak SNR = 5 the GC and NGC correlation types are beginning to properly estimate the trend of increasing shift magnitudes in the positive x-axis direction, but are not providing accurate estimates of the true shift magnitudes. The OC algorithm fails completely for this particular test image and all three SNR values considered. Additionally, the root-mean-squared error (RMSE) of the shift estimates was calculated for each type of correlation. These RMSE values were then averaged over all shift values considered, and these results are shown in Fig. 3 providing a means to compare registration algorithm performance as a function of SNR.



Peak Image SNR = 2 Peak Image SNR = 3 Peak Image SNR = 5

Fig. 2. Expected value of translation estimates produced by four correlation types as a function of true translation

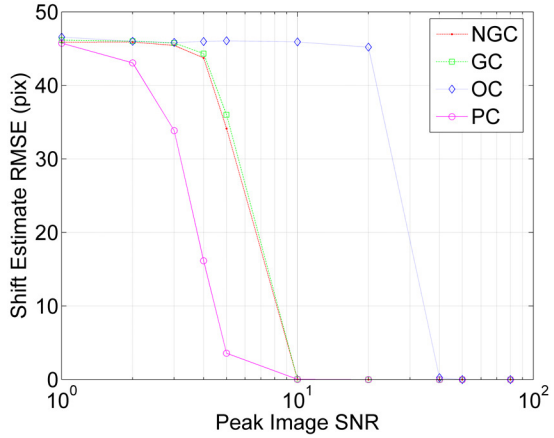


Fig. 3. RMSE of shift estimates produced by four correlation types, averaged over all shifts considered, as a function of peak image SNR

The result of Fig. 3 show that registration error for PC, given the simple synthetic test image, begins to quickly drop for peak image SNRs of approximately 2 and greater, while the RMS errors for GC and NGC do not begin to decrease until SNR increases to approximately 4. The errors for OC remain very large until the peak image SNR is greater than 20. Note that the study producing Figs. 1-3 was repeated for a natural and more complex image. The only significant difference for those results (not presented here) was that good registration performance was achievable at lower peak image SNRs. The relative performance of the four correlation types was the same as shown in Fig. 3.

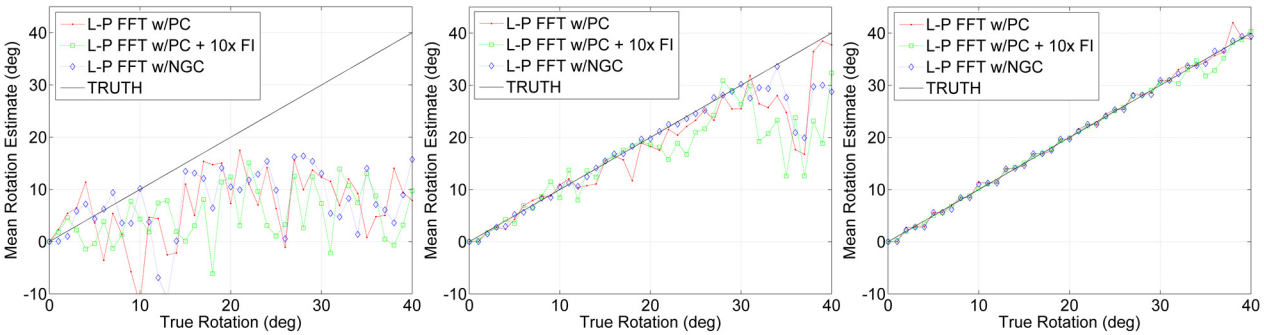
ROTATION ESTIMATION PERFORMANCE FOR THREE L-P FFT ALGORITHM VARIATIONS

Based on the results of Figs. 1-3 we explored variations of L-P FFT registration based on both PC and NGC. The next set of results is for a set of images with simulated rotation (using bilinear interpolation), but for no translations or scale change. The baseline test image is a 128×128 pixel cropped version of Lenna. This image was then rotated clockwise around image center by an integer number of degrees ranging from $\theta_0 = [0,\ldots,40]$ resulting in a 41 image sequence. Three example images are shown below in Fig. 4 for $\theta_0 = 0$ and peak image SNR values of 5, 10, and 20.

Peak Image SNR = 5 Peak Image SNR = 10 Peak Image SNR = 20

Fig. 4. Cropped Lenna image used for testing rotation registration algorithms (shown for three peak SNR values)

Rotation between the baseline image and each rotated image was calculated using three separate versions of the L-P FFT algorithm based on either PC, NGC, or PC with 10× Fourier interpolation, for 250 realizations of noise, and for peak image SNR levels of [1,2,3,4,,5,10,20,40,80,∞]. The expected value of the rotation estimates were calculated as a function of true rotation, and these results are plotted in Fig. 5 for peak image SNR values of 5, 10, and 20.



Peak Image SNR = 5 Peak Image SNR = 10 Peak Image SNR = 20

Fig. 5. Expected value of rotation estimates produced by three versions of the L-P FFT registration algorithm as a function of true rotation

The results of Fig. 5 suggest that L-P FFT rotation registration based on NGC provides somewhat improved performance as compared to L-P FFT registration based on PC. The result presented in Fig. 5 for a peak image SNR of 10 shows noticeably less variation in the mean rotation estimates for L-P FFT registration based on NGC than for the two versions based on PC. Also note that at a peak image SNR of 5 all three algorithms fail to provide useful estimates, while at SNR = 20 all algorithms are providing good rotation estimates for the given image and range of rotations considered. However, even at SNR =20 the L-P FFT NGC algorithm is providing more stable estimates for rotations angles in the range of θ_0 = 30-40 degrees. Additionally, the RMSE of the rotation estimates was calculated for the three L-P FFT algorithm variations studied. These RMSE values were then averaged over all rotation values considered, and these results are shown in Fig. 6 providing a means to compare registration algorithm performance as a function of SNR.

The result of Fig. 6 shows that overall rotation registration error for the L-P FFT algorithm using NGC is noticeably less than for L-P FFT registration using PC. The rotation estimate RMSE for L-P FFT using NGC is less than the PC-based variations for peak image SNR values in the range 3 to nearly 40. All three algorithms perform poorly for peak SNR < 3. For peak image SNR values greater than approximately 40 the L-P FFT algorithm using PC with 10× Fourier interpolation begins to outperform the algorithm using NGC. At this point, for the given image, the higher SNR values allow Fourier interpolation to provide useful information in refining the location of the peak in the PC surface. Note that this same set of tests was also attempted for the same synthetic image of a 13x13 pixel square used for translation estimation testing. However, all three L-P FFT algorithm variations failed to provide meaningful estimates for this very simple test case. Apparently, this simple image of a square does not contain enough diversity in the angular direction to support rotation registration.

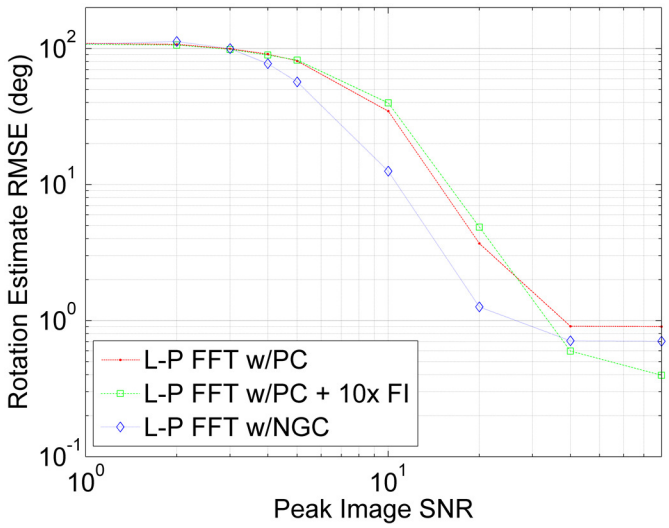
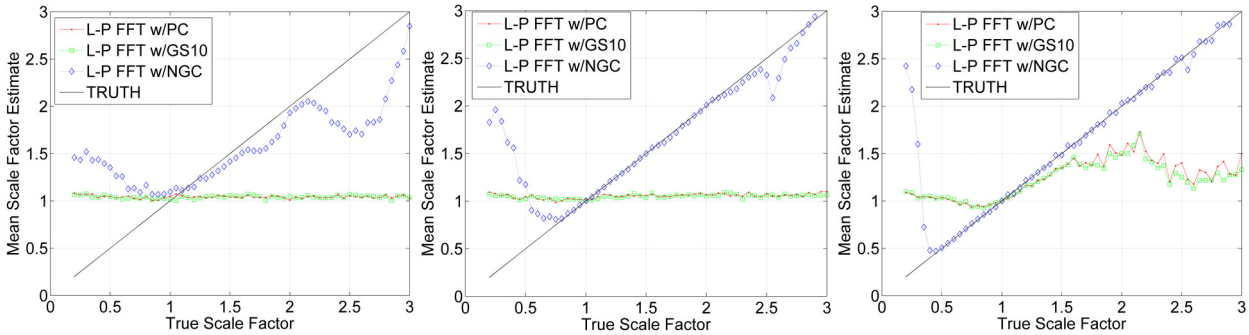


Fig. 6. RMSE of rotation estimates produced by three L-P FFT algorithm variations, averaged over all rotations considered, as a function of peak image SNR

SCALE ESTIMATION PERFORMANCE FOR THREE L-P FFT ALGORITHM VARIATIONS

Based on the results of Figs. 5 and 6 we continued to explore variations of L-P FFT registration based on both PC and NGC. The next set of results is for a set of synthetic images with simulated scale change, but for no translations or rotation. The baseline test image is a 20×20 pixel matrix of ones zero-padded to 128×128 pixels. A set of test images was created by varying the size of the matrix of ones by an integer number of pixels ranging from 4 to 60 resulting in a 57 image sequence. The scale factor between each image in the sequence and the baseline image was then estimated using three separate versions of the L-P FFT algorithm based on either PC, NGC, or PC with 10× Fourier interpolation, for 350 realizations of noise, and for peak image SNR levels of [1,2,3,4,,5,10,20,40,80,120,∞]. The expected value of the scale factor estimates were calculated as a function of true scale factor, and these results are plotted in Fig. 7 for peak image SNR values of 2, 4, and 10.



Peak Image SNR = 2 Peak Image SNR = 4 Peak Image SNR = 10

Fig. 7. Expected value of scale factor estimates produced by three versions of the L-P FFT registration algorithm as a function of true scale factor

The results of Fig. 7 indicate that for this simple test case the L-P FFT algorithm using NGC clearly outperforms the two L-P FFT variations based on PC. For a peak image SNR of 2 the L-P FFT algorithm using NGC is beginning to provide meaningful estimates in the range s = 1 to 2. For a peak image SNR of 4 the algorithm is yielding very good estimates, in terms of the expected value, in the range s = 0.8 to 3 with the exception of several points near 2.6. The L-P FFT algorithm variations based on PC fail completely for SNR values of 2 and 4. Note that the algorithms were written to provide a default scale estimate near 1 for the case of multiple identical peaks in the PC surface - this is the probable cause for the PC-based estimates being near to one regardless of true scale factor for SNR values of 2 and 4. Also note that at peak SNR = 10 the PC-based algorithms are beginning to provide meaningful estimates in the range s = 1 to 1.7. Additionally, the RMSE of the scale factor estimates was calculated for each algorithm. These RMSE values were then averaged over all scale factors considered, and these results are shown in Fig. 8 providing a means to compare registration algorithm performance as a function of SNR.

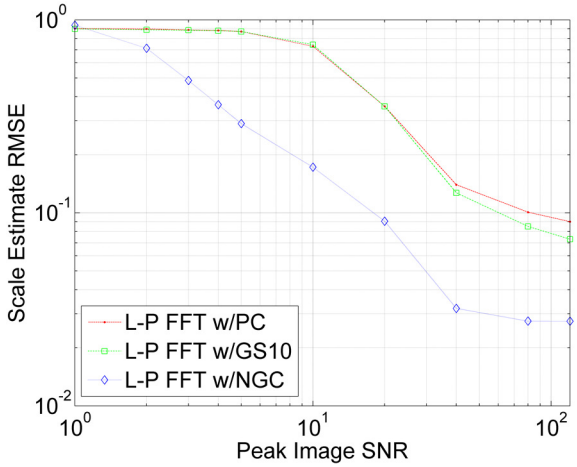
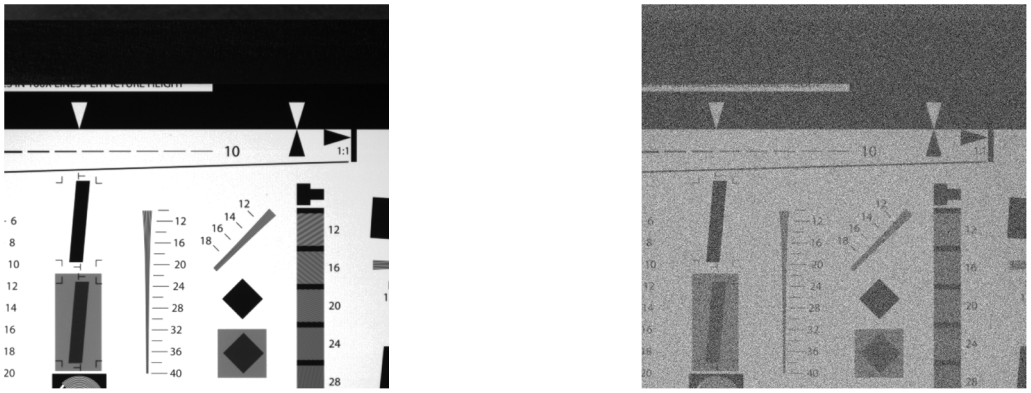


Fig. 8. RMSE of scale factor estimates produced by three L-P FFT algorithm variations, averaged over all scale factors considered, as a function of peak image SNR

The result of Fig. 8 shows that for this simple test case the scale factor estimation performance of the L-P FFT algorithm using NGC is superior to that for L-P FFT registration using PC. The scale estimate RMSE for L-P FFT using NGC is less than the PC-based variations for all peak image SNR values studied greater than 1. All three algorithms perform poorly for peak SNR = 1. The RMSE for the NGC-based algorithm begins to quickly decrease for peak image SNR values greater than 1, while RMSE for the PC-based algorithms does not begin to noticeably decrease until SNR becomes larger than 5. For SNR values of 40, 80, and 120 the L-P FFT algorithm using PC with 10× Fourier interpolation begins to slightly outperform the algorithm using pixel-level PC. At this point, for the given image, the higher SNR values allow Fourier interpolation to provide useful information for refining the location of the peak in the PC surface. However, Fourier interpolated-PC is still unable to provide performance comparable to that for NGC operating at pixel-level – even at the highest SNR levels studied.

TRANSLATION ESTIMATION PERFORMANCE FOR A HYBRID ALGORITHM

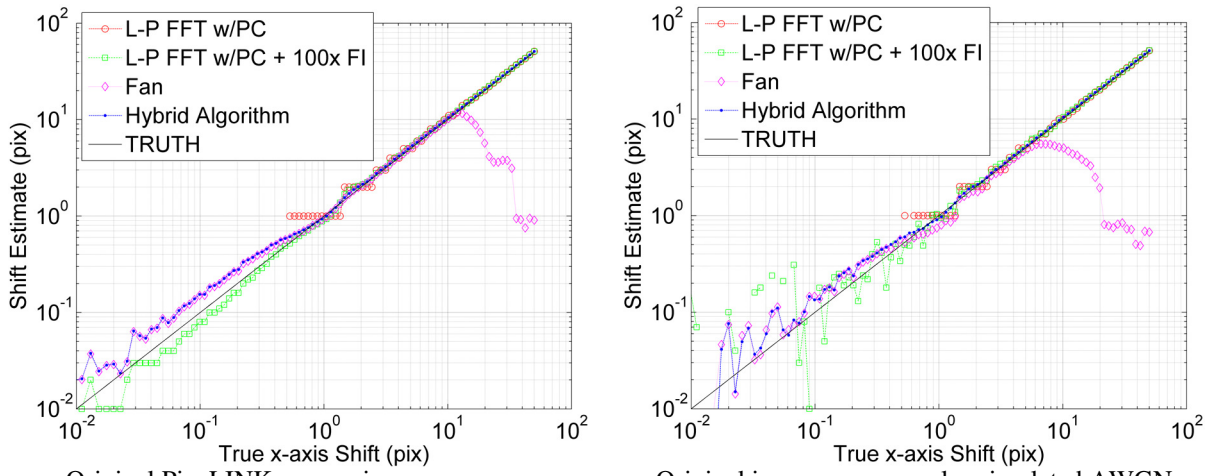
The overriding goal of this paper is to explore robust registration based on combining the Fourier and spatial domain techniques. To that end, the next result is a comparison of translation estimation performance of the spatial domain technique of Fan [10], two variations of L-P FFT registration based on either PC [3] or PC with 100× Fourier interpolation [6-7], and a hybrid algorithm using the result of log-polar FFT registration (based on PC with 100× Fourier interpolation) as an initial guess for the spatial domain algorithm of Fan. Translation estimates were calculated using these four algorithms for a sequence of images captured in the laboratory using a PixeLINK CMOS machine vision camera mounted on several computer-controlled micro-stepping motion stages. The test object is an ISO 12233 resolution chart, and our laboratory setup was previously described in greater detail [6,11]. The camera was shifted perpendicular to line-of-sight along the *x*-axis based on a set of nearly log-spaced positions, resulting in a 101 image sequence; the minimum and maximum calibrated shifts are 0.001 and 49.974 pixels, respectively. The center 512×512 pixels were extracted from each 1024×1280 camera frame prior to algorithm testing, and translation estimates were calculated for both the original 8-bit bitmap images recorded using the PixeLINK image capture software, as well as for this same image set corrupted by additional simulated AWGN defined by a peak image SNR of 4. The first image in the *x*-axis shift sequence is shown below in Fig. 9, both the original frame (left) and the original plus additional simulated AWGN (right).



Laboratory image Laboratory image plus simulated AWGN

Figure 9. The first PixeLINK camera image in the 101 image *x*-axis shift sequence captured in the laboratory. The original frame is on the left, and the original plus simulated AWGN is on the right for peak image SNR = 4.

Translation between the first image and each image in the sequence was estimated using each of the four algorithms identified above, and the results are presented in Fig. 10 as a function of true (calibrated) shift. Results are shown on the left for the original image sequence and on the right for the case of additional simulated AWGN defined by a peak image SNR of 4.



Original PixeLINK camera image sequence Original image sequence plus simulated AWGN

Figure 10. Translation estimates produced by four registration algorithms, including a hybrid algorithm combining the L-P FFT and spatial domain techniques. The true (calibrated) shifts are plotted as a solid black line.

The results of Fig. 10 indicate that for this well-controlled translation-only test case, the hybrid algorithm has successfully merged the performance of the L-P FFT and spatial domain techniques. As expected, the spatial domain technique of Fan alone provides very good performance for small shifts but has a limited dynamic range. For the original PixeLINK camera image sequence, the Fan algorithm is unable to accurately estimate shifts larger than approximately 10 pixels. For the case of the original image sequence corrupted by additional simulated AWGN, the Fan algorithm is unable to accurately estimate shifts larger than approximately 6 pixels. However, in both test cases the hybrid algorithm closely follows the performance of the Fan algorithm at small shifts, and that of the L-P FFT techniques at large shifts. For cases where the initial guess represents a relatively large motion, success of the spatial domain portion of the hybrid algorithm is heavily reliant on proper identification of valid pixels. For this particular test case, the hybrid algorithm failed for large shifts prior to implementing a Boolean mask to avoid use of invalid pixels corresponding to those parts of the object not imaged by both frames. Also note the results of Fig. 10 show that for the original image sequence, the hybrid algorithm and L-P FFT algorithm based on 100× Fourier interpolation provide nearly equal performance. The spatial domain technique requires less memory, although the efficient Fourier interpolated PC algorithm of Guizar-Sicairos [7] mitigates the memory issue to a great extent. For the case of additional simulated AWGN, however, the spatial domain algorithm provides noticeably more stable estimates of sub-pixel shifts. This observation is consistent with other simulations, not reported here, that showed spatial domain techniques providing more robust performance at low SNR levels, and an overall more graceful degradation as SNR decreases.

OPTICAL FLOW INVESTIGATIONS

In space object imagery, resident space objects (RSOs) often rotate, spin, and move in other complicated ways, in addition to simple global translations and rotation. Therefore, we have also investigated optical flow techniques to characterize these more complex motions. In order to deal with expected object motion and changing lighting conditions within a series of frames, optical flow techniques are explored to assist with image registration. Although the idea of using optical flow for feature tracking in a series of images has been explored previously, the use of optical flow for multi-frame super-resolution is still a relatively new concept [13]. In previous work we showed $1/10^{th}$ of a pixel registration with linear shifts [6]. In this paper we show investigations of imagery of a rotating object, where similar accuracies are achieved. Another characteristic of real space objects is that because their reflection characteristics are geometry dependent, the dynamic range of a set of successive images of a moving object can vary by several orders of magnitude due to glinting. Understanding how such dynamic range variation degrades performance is also of interest.

With even a basic understanding of optical flow and its implementation, one notices that the motion captured at a specific location in a series of frames is estimable as a set of linear translations. For example, rotation of an imaged 3-D object feature will appear to be only a slightly translated from frame to frame, if the object is imaged over a time that is short and at a high enough rate of image capture. The trivial limit of this is when the frame rate is fast

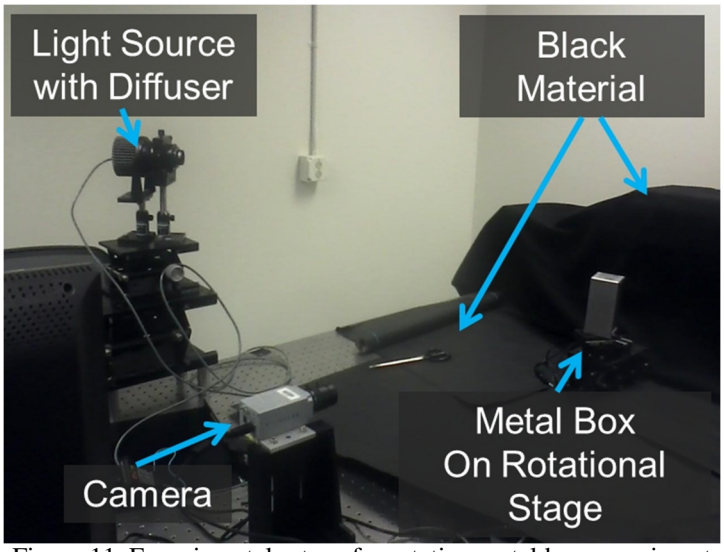


Figure 11. Experimental set-up for rotating metal box experiment

enough that the object appears to barely move at all between successive frames. The conditions that i) the object is imaged over a short time interval, and ii) that the image capture rate is fast enough such that motion of the object between successive frames is small, are together called the adiabatic condition. Mathematically, the optical flow equations are valid only when the adiabatic condition applies. Part of our research is to understand when the results of an optical flow calculation break down, how that breakdown affects accuracy of results, and if some of the standard limitations may be overcome.

Controlled laboratory data that displays complicated motion and changing lighting conditions was desired; therefore we collected laboratory imagery of a rotating metal box. A picture of the experimental setup is shown in Fig. 11. In the setup, an LED lamp backing a ground glass diffuser provides a light source. The imaged object is a 3.87 cm metal BUD box, positioned in the center of a rotational stage. Black cloth was arranged behind the metal box and black aluminum foil on the optics table to minimize stray light and cover other objects in the imagery. A PixeLINK camera with 25 mm focal length lens (f/4 aperture) imaged the rotating box from ~74 cm away – the exact distance between camera and object depends on whether a face or an edge of the box faces the camera.

An extensive, well-controlled set of 332 images was collected. An example image sequence is shown in Fig. 12. The finest and coarsest rotation steps allowed by this particular hardware setup are 0.099981° and 0.233°, respectively. However, one can skip several or even dozens of frames to create a sequence of fewer frames that has up to 355° of total motion.

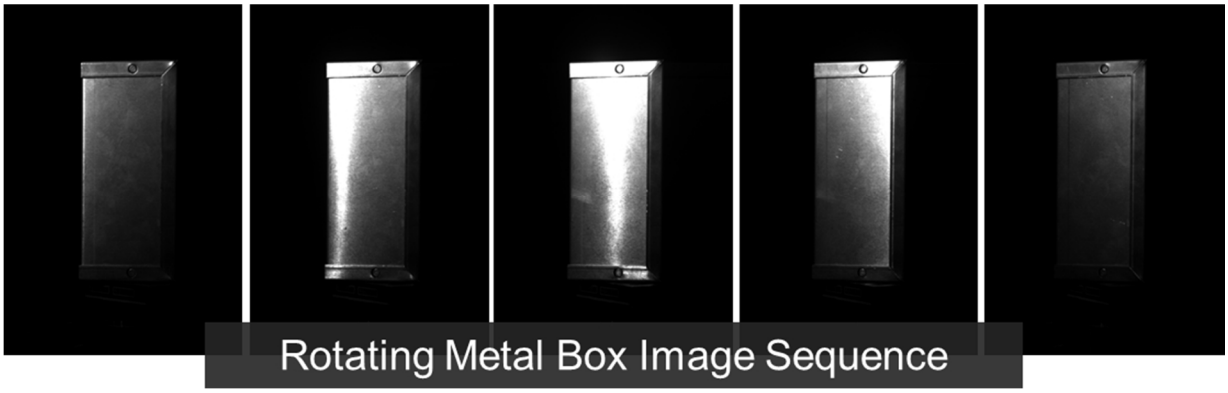


Figure 12. Image sequence of rotating metal box collected in the laboratory. As the box rotates relative to the camera and light source, glints are observed.

The specific optical flow algorithm studied is the phase-based technique of Gautama and Van Hulle (GVH) published in 2002 [14]. In addition to the technique of GVH, we investigated a few other approaches for optical flow calculation including those of Lucas-Kanade and Horn and Schunck. Preliminarily, we also investigated the use of the technique of Brox, et al. [15]. In general, these algorithms all require fine tuning, and from that perspective we made the most progress with that of GVH, and found it to be more robust than Horn and Schunck and Lucas-Kanade for a limited set of test imagery. However, we have not ruled out the usage of those algorithms or others in future experiments. One goal of future work is to understand the limitations and strengths of these diverse algorithmic approaches, especially when applied to diverse and realistic space imagery.

With this well controlled data we began investigations into the GVH algorithm for the motion of a rotating 3-D box that exhibits glints. When the box is edge on, as shown in Fig. 13, the projected velocity varies from a maximum at the nearest edge to a projected value of zero at the sides.

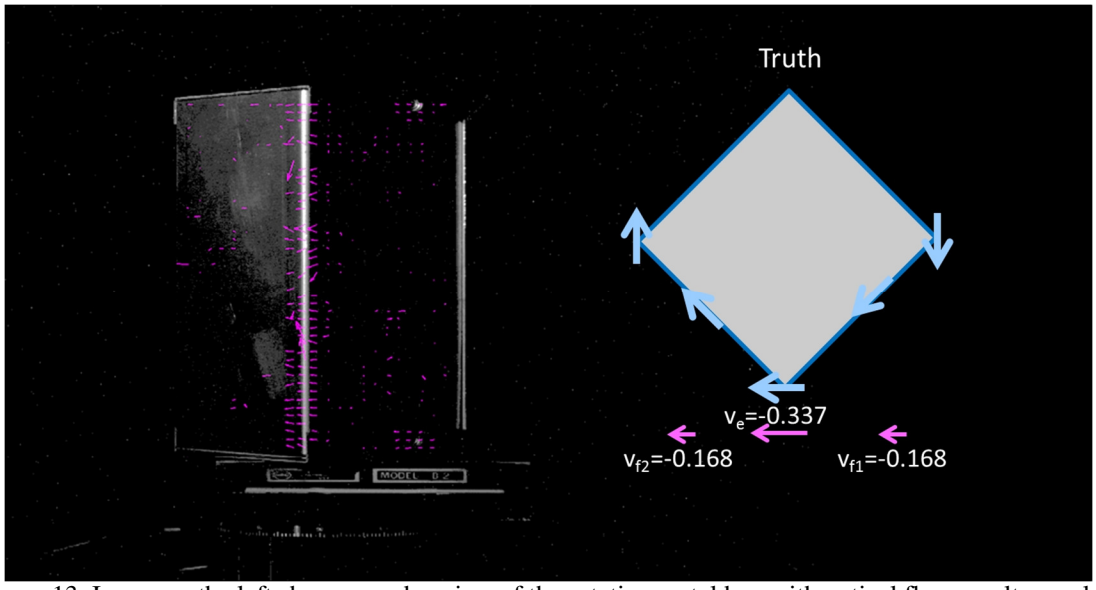


Figure 13. Image on the left shows an edge view of the rotating metal box with optical flow results overlaid (magenta arrows). Diagram on the right shows a top view of the expected velocity across the image: blue arrows represent 3D-box velocity and magenta arrows represent velocity projected into image plane. Note the metal box image contrast and brightness are increased due to high dynamic range across image; raw data was processed. The box is rotating on a stage at a rate -0.0228°/frame or relative velocity at the box edge of -0.337 pixels/frame.

Results from the GVH optical flow algorithm are $\hat{v}_x = -0.31$ pixels/frame on edge and $\hat{v}_x = -0.14$ pixels/frame midway across the face of the box, within 0.028 pixels/frame of what we estimate to be the truth. Although these results suggest accuracies to better than 1/30th of a pixel, estimates of errors in our ability to autonomously determine the velocity at various locations across the image suggest accuracies on the order of 1/5th of a pixel. These results are promising but more investigations are required to understand when the adiabatic condition begins to fail, and how to autonomously determine frame-to-frame motion.

5. SUMMARY AND CONCLUSIONS

This paper presents results from three distinct, but related, studies, all aimed at robust estimation of image motions primarily for application to image enhancement and high dynamic range imaging. Our approach to developing a robust *global* registration method is based on a hybrid algorithm combining the L-P FFT and spatial domain techniques. We also present results from an investigation of methods for calculating *optical flow*; one avenue to estimate and describe more complicated, non-affine motion scenarios.

The first set of results is motivated by considering the choice of correlation type at each relevant step in the L-P FFT algorithm as a potential means to increase overall robustness. Simulation results are presented for four correlation types: phase-only correlation (PC), gradient correlation (GC), normalized gradient correlation (NGC), and

orientation correlation (OC). While the 1996 paper of Reddy describes a version of the L-P FFT algorithm based entirely on PC, the more recent work of Tzimiropoulos recommends an algorithm based entirely on NGC. However, our results strongly suggest that a combination of PC and NGC is actually a better approach than using either PC or NGC alone. Within the L-P FFT construct, rotation and scale estimates are derived from correlation in the log-polar Fourier magnitude domain, while translation estimates are derived from correlation in the image domain. We studied rotation estimation performance using a simulated sequence of Lenna images, corrupted by various levels of AWGN. Translation and scale estimation performance was studied using simulated sequences of a simple synthetic image of a square, also corrupted by various levels of AWGN. In terms of parameter-averaged RMSE, rotation and scale estimation performance was significantly better for the L-P FFT algorithm based on NGC as compared to the PC-based version. On the other hand, parameter-averaged RMSE results for translation estimation were significantly better for the L-P FFT algorithm based on PC as compared to the NGC-based version. The second main result of this paper is successful demonstration of a hybrid algorithm using the result of log-polar FFT registration (based on Fourier interpolated PC) as an initial guess for the spatial domain algorithm of Fan. This result is based on translation estimation analysis of a set of images of an ISO 12233 resolution chart captured in the laboratory using a CMOS machine vision visible camera. For the original image sequence, results demonstrate nearly equal performance for the hybrid algorithm and for the L-P FFT algorithm based on PC with 100× Fourier interpolation. For the case of additional simulated AWGN, however, the spatial domain algorithm provides noticeably more stable estimates of sub-pixel shifts. This observation is also consistent with other simulations (not reported here) that showed spatial domain techniques providing more robust performance at low SNR levels, and an overall more graceful degradation as SNR decreases. Finally, we presented results from our optical flow investigations based on a sequence of laboratory images of a spinning metal box. Based on our studies to date, the optical flow technique of Gautama and Van Hulle is more robust than that of Horn and Schunck. Analysis of the spinning metal box image sequence yields sub-pixel motion estimation on the order of $1/5^{th}$ pixel accuracy.

Plans for future work include a more detailed analysis of the performance limits of PC and NGC in both the image and log-polar Fourier magnitude domains. This may include calculating Cramér-Rao lower bounds for estimation of scale, rotation, and translation parameters. Our ultimate goal is pursuit of a robust registration and multi-frame image enhancement algorithm leveraging both global and feature-based concepts. To this end, future results will also be considered from the point-of-view of a more useful metric for defining signal in regards to image registration, such as a weighted sum of the squared Fourier magnitude components. Application of the optical flow techniques to image enhancement also requires additional development and study. In addition, we plan to extend the image domain methods of Keren and Fan to provide scale estimation in addition to rotation and translation, and will also consider creating an efficient Fourier interpolated version of NGC based on the PC work of Guizar-Sicairos, et al. Future work will also further consider the optimal combination of L-P FFT and spatial domain techniques, to include a comparison of our hybrid algorithm with a multi-resolution implementation of a spatial domain method. We will explore and refine application of the iterative image domain techniques to the correlation of log-polar mapped Fourier moduli in the L-P FFT algorithm. Improved accuracy of the scale and rotation estimates at this point in the algorithm will propagate through to also improve subsequent estimates of *x*- and *y*-axis translations. Finally, future algorithm performance evaluations will consider dynamic lighting conditions and limitations of real cameras beyond simulated AWGN.

6. REFERENCES

1. Brown, L.G., “A survey of image registration techniques,” *ACM Comput. Surv.* **24**(4), 325-376 (1992).
2. Zitová, B. and Flusser, J., “Image registration methods: a survey,” *Image and Vision Computing* **21**(11), 977-1000 (2003).
3. Reddy, B.S. and Chatterji, B.N., “An FFT-based technique for translation, rotation, and scale-invariant image registration,” *IEEE Trans. on Image Proc.* **5**(8), 1266-1271 (1996).
4. Tzimiropoulos, G. and Stathaki, T., "Robust FFT-based scale-invariant image registration," *4th SEAS DTC Technical Conference* (2009).
5. Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., and Stathaki, T., "Robust FFT-based scale-invariant image registration with image gradients," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **32**(10), 1899-1906 (2010).
6. Crabtree, P.N. and Murray-Krezan, J., “Geometric super-resolution via log-polar FFT image registration and variable pixel linear reconstruction,” *SPIE Proc. 8165, Unconventional Imaging, Wavefront Sensing, and Adaptive Coded Aperture Imaging and Non-Imaging Sensor Systems*, 81650O (September 13, 2011).

7. Guizar-Sicairos, M., Thurman, S.T., and Fienup, J.R., "Efficient subpixel image registration algorithms," *Opt. Lett.* **38**(2), 156-158 (January 15, 2008).
8. Hardie, R.C., Barnard, K.J., Bognar, J.G., Armstrong, E.E., and Watson, E.A., "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Opt. Eng.* **37**(1), 247-260 (January 1, 1998).
9. Keren, D., Peleg, S., and Brada, R., "Image sequence enhancement using sub-pixel displacements," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 742-746 (1988).
10. Fan, C., Gong, J., Zhu, J., and Zhang, L., "An improvement approach based on Keren sub-pixel registration method," *Proc. IEEE 8th International Conference on Signal Processing*, 16-20 (2006).
11. Crabtree, P.N., Dao, P.D., and Picard, R.H., "Experimental investigation of the performance of image registration and de-aliasing algorithms," *Advanced Maui Optical and Space Surveillance Technologies Conference*, Maui, HI (2009).
12. ISO 12233:2000(E), Photography – Electronic still-picture cameras – Resolution measurements.
13. Kanaev, A. and Miller, C.W., "Confidence measures of optical flow for multi-frame image reconstruction," *OSA Conference on Computational Optical Sensing and Imaging (COSI)*, Monterey, California (June 25, 2012).
14. Gautama, T. and Van Hulle, M.M., "A Phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Trans. Neural Networks* **13**(5), 1127-1136 (2002).
15. Brox, T., Bruhn, A., Papenberg, N., and Weickert, J., "High accuracy optical flow estimation based on a theory for warping," *Proc. 8th European Conference on Computer Vision, vol. 4*, 25-36, Prague Czech Republic (May 2004).